I/we claim:

1. A method of verifying and controlling assays for the analysis of nucleic acid variations by means of statistical process control, characterized in that variables of each experiment are monitored by measuring deviations of said variables from a reference data set and wherein said experiments or batches thereof are indicated as unsuitable for further interpretation if they exceed predetermined limits.

2. A method according to claim 1 when said nucleic acid variations are cytosine methylation variations.

3. A method according to claims 1 and 2 wherein said statistical process control is taken from the group comprising multivariate statistical process control and univariate statistical process control.

4. A method according to claims 1 to 3 comprising the following steps
a) defining a reference data set
b) defining a test data set
c) determining the statistical distance between the reference data set and test data set or elements or subsets thereof
d) identifying individual elements or subsets of the test dataset which have a statistical distance larger than that of a predetermined value.

5. The method according to claim 4, further comprising in step b)
reducing the data dimensionality of the reference and test data set by means of robust embedding of the values into a lower dimensional representation.

6. The method according to claim 5 wherein step b) is carried out by calculating the embedding space using one or both of the reference and the test data sets.

7. The method according to one of claims 4 to 6 further comprising,
e) further investigating said identified elements or subsets of the test dataset to determine the contribution of individual variables to the determined statistical distance.

8. The method according to one of claims 4 to 7 further comprising,
e) excluding said identified experiments or batches thereof from further analysis.

9. The method of claim 4 wherein in step d) said statistical distance is calculated by means of one or more methods taken from the group consisting the Hotelling's T^2 distance between a single test measurement vector and the reference data set, the Hotelling'-T^2 distance between a subset of the test data set and the reference data set, the distance between the covariance matrices of a subset of the test data set and the covariance matrix of the reference set, percentiles of the empirical distribution of the reference data set and percentiles of a kernel density estimate of the distribution of the reference data set, distance from the hyperplane of a nu-SVM, estimating the support of the distribution of the reference data set.

10. The method according to one of claims claim 5 and 6 wherein the data dimensionality reduction is carried out by means of principle component analysis.

11. The method according to one of claims claim 5, 6 and 10 wherein the data dimensionality reduction step comprises the following steps
i) Projecting the data set by means of robust principle component analysis
ii) Removing outliers from the data set according to their statistical distances calculated by means of one or more methods taken from the group consisting

of: Hotelling's T^2 distance; percentiles of the empirical distribution of the reference data set; Percentiles of a kernel density estimate of the distribution of the reference data set and distance from the hyperplane of a nu-SVM, estimating the support of the distribution of the reference data set.

iii) Calculating the embedding projection by standard principle component analysis and projecting the cleared or the complete data set onto this basis vector system.

12. The method according to one of claims 4 to 11 wherein at least one of the variables measured in steps a) and b) is determined according to the methylation state of the nucleic acids.

13. The method according to one of claims 4 to 11 wherein at least one of the variables measured in step a) and b) is determined by the environment used to conduct the assay.

14. The method according to one of claims 4 to 11 wherein said data sets comprises one or more variables selected from the group comprising mean background/baseline values; scatter of the background/baseline values; scatter of the foreground values, geometrical properties of the array, percentiles of background values of each spot and positive and negative assay control measures.

15. A method according to one of claims 4 to 14 wherein the reference data set is the complete series of experiments being analysed. (make it explicit in the description that the test set can be a subset of the reference data set.)

16. A method according to one of claims 4 to 14 wherein the reference data set is derived from experiments carried out separately to those of the test data set.

17. A method according to one of claims 4 to 14 wherein the reference data set is derived from a set of experiments wherein the value of each variable of each experiment is either within a predetermined limit or optimally controlled.

18. A method according to one of claims 4 to 17 further comprising the generation of a document comprising said elements or subsets of the test data determined according to step d) of claim 4.

19. A method according to claim 18 wherein said document further comprises the contribution of individual variables to the determined statistical distance.

20. A method according to claims 18 and 19 wherein said document is stored on a computer readable format.

21. A method according to one of claims 1 to 20 wherein said method is implemented by means of a computer.

22. A computer program product for the verifying and controlling assays for the analysis of nucleic acid variations comprising

a) a computer code that receives as input a reference data set

b) a computer code that receives as input a test data set

c) a computer code that determines the statistical distance between the reference data set and test data set or elements or subsets thereof

d) a computer code that identifies individual elements or subsets of the test dataset which have a statistical distance larger than that of a predetermined value

e) a computer readable medium that stores the computer code.

23. The computer program product of claim 22 further comprising

f) a computer code that reduces the data dimensionality of the reference and test data set by means of robust embedding of the values into a lower

dimensional representation.

24. The computer program product of claim 22 characterised in that the embedding space is calculated using one or both of the reference and the test data sets.

25. The computer program product of claims 22 to 24 further comprising,
g) a computer code that investigates said identified elements or subsets of the test dataset to determine the contribution of individual variables to the determined statistical distance.

26. The computer program product of claims 22 to 25 wherein said statistical distance is calculated by means of one or more methods taken from the group consisting the Hotelling's T^2 distance between a single test measurement vector and the reference data set, the Hotelling'-T^2 distance between a subset of the test data set and the reference data set, the distance between the covariance matrices of a subset of the test data set and the covariance matrix of the reference set, percentiles of the empirical distribution of the reference data set and percentiles of a kernel density estimate of the distribution of the reference data set, distance from the hyperplane of a nu-SVM, estimating the support of the distribution of the reference data set.

27. The computer program product of claims 23 and 24 wherein the data dimensionality reduction is carried out by means of principle component analysis.

28. The computer program product of claims 23, 24 and 27 wherein the data dimensionality reduction step comprises the following steps
i) Projecting the data set by means of robust principle component analysis
ii) Removing outliers from the data set according to their statistical distances calculated by means of one or more methods taken from the group consisting of: Hotelling's T^2 distance; percentiles of the empirical distribution of the

reference data set; Percentiles of a kernel density estimate of the distribution of the reference data set and distance from the hyperplane of a nu-SVM, estimating the support of the distribution of the reference data set.

iii) Calculating the embedding projection by standard principle component analysis and projecting the cleared or the complete data set onto this basis vector system.

29. The computer program product of claims 22 to 28 further comprising a computer code that generates a document comprising said elements or subsets of the test data determined according to step d) of claim 22.